Frequency domain analysis of telephone helpline call data

FREQUENCY DOMAIN ANALYSIS OF TELEPHONE HELPLINE CALL DATA

ALEXANDER GRIGORASH, RAYMOND R. BOND, MAURICE D. MULVENNA
School of Computing, Ulster University, Shore Road, Newtownabbey, BT37 0QB, UK

SIOBHAN O'NEILL, CHERIE ARMOUR, COLETTE RAMSEY
School of Psychology, Ulster University, Cromore Road, Coleraine, BT52 1SA, UK

The paper presents a frequency domain analysis of call data records of a telephone helpline for those seeking mental health and wellbeing support and for those who are in a suicidal crisis. A call data record dataset provided by Samaritans Ireland helpline is used. Fourier series is used to ascertain periodicity in the call volume. The main findings from the paper indicate that strong repetitive intra-day and intra-week patterns are found, while intra-month repetitions are conspicuously absent.

1. Introduction

TELEPHONE helplines provide a significant service that continues to underpin mental wellbeing and suicide prevention efforts [1]. This work is part of a research project that involves analysis of digital telephony data provided by Samaritans Ireland, a charity with a helpline to provide emotional support to anyone in distress or at risk of suicide.

Data were provided for all calls made to Samaritans in Ireland from April 2013 to December 2016, a total of 3.449 million inbound calls. Each data record carries the date-time stamp of the call arrival. The arrival time stamps are resolved to the last second of time. There are no simultaneous arrivals.

Informal observations by Samaritans personnel suggested that certain callers dial in at regular intervals. Inspecting the data revealed apparent peaks and troughs in the call volume, *i.e.* the number of calls, arriving to the call centre (Figure 1).

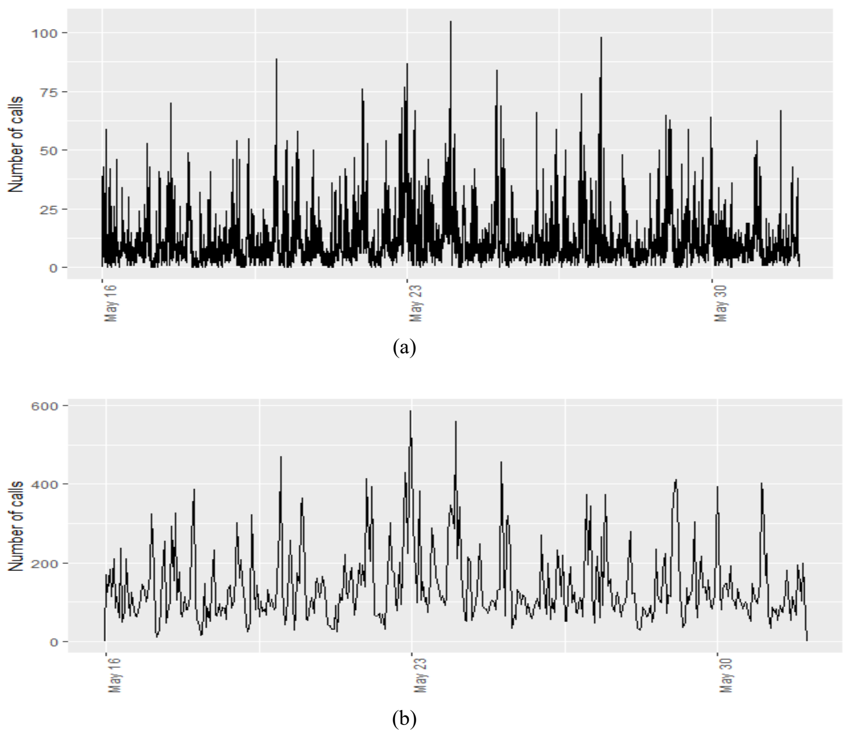

Figure 1. Samaritans call arrival data (two weeks in May 2016) viewed through two different aggregation intervals: (a) 5 minutes, (b) 1-hour.

However, seeing cyclic fluctuations in data does not necessarily mean seeing seasonal, *i.e.*, periodic, regularities, see *e.g.*, [2]. Thus, it became interesting to explore what period lengths (hourly, daily, weekly, monthly, etc.), if any at all, find support in the dataset?

The following referenced, explanatory quote summarizes the intentions and sets out relevant terminology, with our own emphases: “Put simply, a *time-domain* graph shows how a signal changes over time, whereas a *frequency-domain* graph shows how much of the signal lies within each given frequency band over a range of frequencies. … A given function or signal can be converted between the time and frequency domains with a pair of mathematical operators called a *transform*. An example is the *Fourier transform*, which converts the time function into a sum of sine waves of different frequencies, each of which represents a frequency component. The 'spectrum' of frequency components is the frequency domain representation of the signal. The *inverse Fourier transform* converts the frequency domain function back to a time function" [3].

Recent literature surveys [4, 5, 6] show a common approach to converting the point process of incoming calls into a time series. The timeline is split into time buckets, typically 5 minutes or 15 minutes or 30 minutes long. Call arrival times (points) are counted for each bucket. This procedure generates a sequence of call volume values (call counts) versus times, *i.e.*, a time series

Relevant time domain techniques include seasonal moving average, multiplicative seasonal ARIMA and Holt-Winters exponential smoothing [7]. Irregular spikes in call volume are filtered out with machine learning approaches [8] thereby enabling the use of call volume models that expect the data to be reasonably smooth. An older approach relies on statistical analysis alone, that involves judiciously censoring the data set: call arrivals deemed atypical by the researchers are removed from consideration [9].

Surprisingly, sources on frequency domain analysis in the context of call center data are rare. A single Australian conference paper was identified where the authors analyzed 52 weeks of call data records of a police assistance line [10]. They sampled at a resolution of 48 intervals per day, and discovered 11 dominant frequencies, from 1 cycle per week to 4 cycles per day that explained daily and weekly fluctuations of the call volume.

2. Methods

In order to extract frequency components, 2-year-long subset of the data, covering years 2015 and 2016 were used. This time span was chosen because no generic trend that could mask periodic activity was obvious.

In order to convert the call arrival point process into a call count time series, the chosen time span was split into a large number of sampling intervals. The number of calls arriving within each of these intervals (buckets) was counted, and these became the values of the time series. The sampling interval (the bucket size for aggregating calls) was 30 minutes long. This particular size offered the smallest number of sampling intervals at a resolution capable of capturing oscillations as frequent as once per hour.

The fundamental period, i.e., the range of data supporting the analysis, was from 01 January 2015 00:00 to 31 December 2016 23:59, lasting 30,588 sampling intervals exactly. The maximum meaningful frequency detectable at this resolution (Nyquist frequency), equals 17,544 cycles over 2 years, or 731 cycles per month, or 24 cycles per day, or about 168 cycles per week.

During 2015 and 2016, the overall trend in the number of calls per unit of time appeared to be practically static. In order to confirm this, a linear trend line was fitted to the data. The fitted value at the start of the time span was 63.9 calls per sampling interval, the fitted value at the end of the time span was 70.8 calls. This amounts to a very gradual rate of increase, giving about 3.5 calls yearly difference, in line with the perception of a static trend.

R programming language and R Studio were used for data wrangling and to implement the analysis. An open online tutorial by Joao Neto [12]. provided several service functions to simplify working with Fourier transform, which were adopted and used after debugging and making several adjustments.

3. Results

Figures 2-4 depict spectral plots of the Fourier transform. The frequencies of oscillations are plotted against the amplitudes that measure the peak number of calls for each harmonic. The spectral plots are shown after de-trending which amounted to removing the non-oscillatory component that would have shown as frequency of 0 cycles. The strongest set of dominant frequencies corresponds to intra-day oscillations, at 1, 2, 3, 4, 5, 6, 7, 8, and 24 cycles per day (Figure 2).

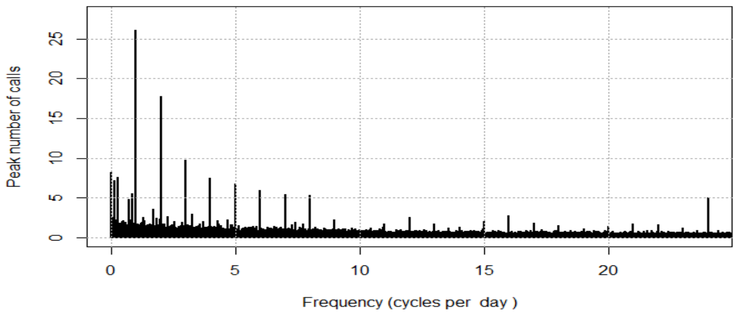

Figure 2. Spectral plot for complete frequency range.

Figure 3 depicts a closer view of the range 0 to 3 cycles per day, expressed using cycles per week. On this plot, the secondary seasonality group of frequencies, associated with intra-week oscillations, is clearly seen. These secondary dominant frequencies are situated at 1 cycle per week, 2, 5 and 6 cycles per week. These frequencies correspond to a natural human inclination to repeat tasks weekly, or twice weekly, or every working day, or every day except *e.g.*, Sunday. The amplitudes of these harmonics range from 5 to 8 calls. The large amplitude at 7 cycles per week shows exactly the same oscillation as the once-a-day amplitude in Figure 2.

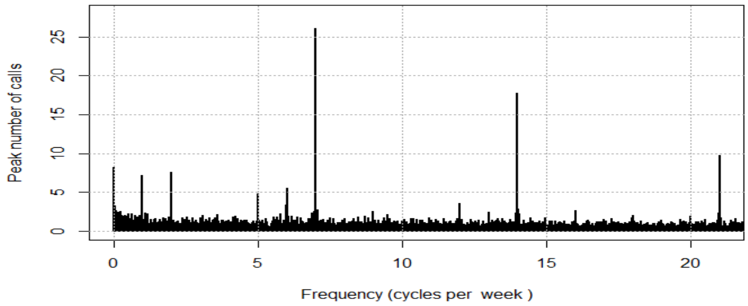

Figure 3. Spectral plot showing the lowest 1/8 of the frequency range, here frequencies are expressed in cycles per week.

Remarkably, few callers seem to regularly call using an ‘every other day’ pattern. The frequency of 0.5 cycles per day, equivalent to 3.5 cycles per week, is indistinguishable from noise.

Further to the right on the plot on Figure 3, the 12 cycles a week frequency can be visually set apart from the noise. However, its strength amounts to about 3.5 calls only, which is between 2 and 2.5 times inferior to the weakest of the previously identified dominant frequencies. Therefore, weaker frequencies in any of the two identified seasonalities are not included.

A closer view at the range 0 to 1 calls per week, expressed as 0 to 4.5 cycles per month, is shown in Figure 4.

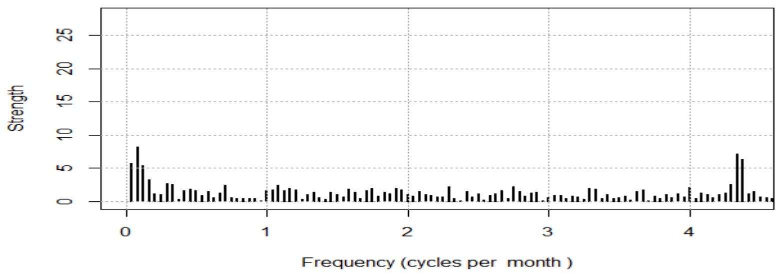

Figure 4. Spectral plot showing the lowest 1/24 of the frequency range, here frequencies are expressed in cycles per month.

On the right of this plot there is an amplitude spike around 4.3 cycles per month, which corresponds to 1 cycle per week. On the left of this plot there is a prominent amplitude of 8 calls situated at 0.2 cycles per month, or about twice a year. The same amplitude is visually shown at near 0 cycles per week on the plot Figure 3. Other frequencies from monthly and quarterly ranges are lacking prominence.

4. Discussion

The volume of calls to Samaritans exhibits strong intra-day and intra-week repetitive patterns, while intra-month repetitions are absent. This double seasonality effect was described in the literature [7, 5]. It appears that the double seasonality is part of the nature of a telephone helpline *per se*, regardless the applied area: mental health and wellbeing helplines, banking [7], sales [8], law enforcement [10].

The once-an-hour, or 24 cycles per day, frequency was part of our intra-day seasonality pattern. As this frequency sits right at the limit of our current resolution, it would be worthwhile re-sampling the time series of calls at a higher rate and re-running the frequency analysis to see if any higher frequencies contribute to the pattern.

The harmonics supported by the data have amplitudes in the range between 4.9 and 27 calls per sampling interval. This makes the seasonal components comparable in magnitude to the linear trend which ranged between 64 and 71 calls per sampling interval.

Future work may involve transforming data in the frequency domain back into the time domain, allowing for call volume modelling and forecasting. However, in this first instance, the primary interest was to test whether periodic activity is present. The answer to that turned out to be a definitive yes.

Acknowledgments

Financial support for this research was provided by Samaritans Ireland with support from Ireland’s National Office for Suicide Prevention.

References

1. S. Howe, B. Meakin, and F. Islam-Barrett. (2014). Helplines at the Frontline of Health and Social Care. *Helplines*. Available: https://www.helplines.org/wp-content/uploads/2014/12/Helplines-at-the-Frontline-of-Health-and-Social-Care.pdf
2. R. J. Hyndman. (14 December 2011), Cyclic and seasonal time series, *Hyndsight blog*. Available: https://robjhyndman.com/hyndsight/cyclicts/

3. Frequency domain, *Wikipedia*. (Accessed 1 Dec 2017) Available: https://en.wikipedia.org/wiki/Frequency_domain
4. R. Ibrahim, H. Ye, P. L'Ecuyer and H. Shen, *Int. J. Forecast.* **32(3)**: 865 – 874 (2016).
5. N. Meade and T. Islam, *Int. J. Forecast.* **31(4)**: 1105–1126 (2015).
6. J. Taylor, *Manage. Sci.* **54(2)**: 253–265 (2008).
7. D. Barrow, J. Bus. Res. **69(12)**: 6088–6096 (2016).
8. D. Barrow and N. Kourentzes, *Eur. J. Operat. Res.* **264(3)**: 967-977 (2016).
9. L. Brown, et al, *J. Am. Statist. Assoc.* **100(469)**: 36–50 (2005).
10. B. Lewis, R. Herbert, and R. D. Bell, *Proc. Int. Cong. Model. Simul. (MODSIM03)*, Townsville, Australia, 1281–1286 (2003). Available: https://www.mssanz.org.au/MODSIM03/Volume_03/B10/06_Lewis.pdf
11. J. Neto. (March 2013). Fourier Transform: A R Tutorial. Available: http://www.di.fc.ul.pt/~jpn/r/fourier/fourier.html